

Phylogenetic shadowing:

A Novel Tool To Identify Primate-Specific Regulatory Elements

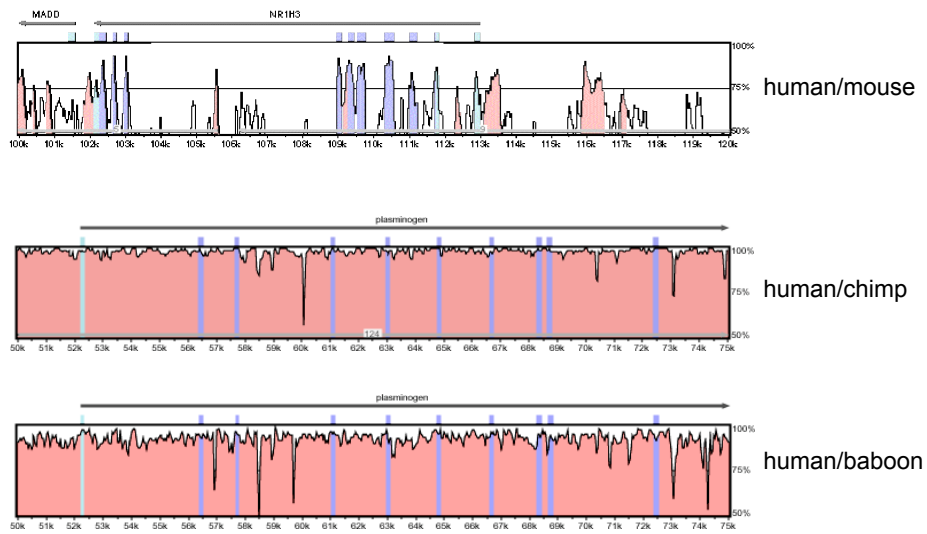
Approach to the Identification of Primate-Specific Regulatory Elements



VS

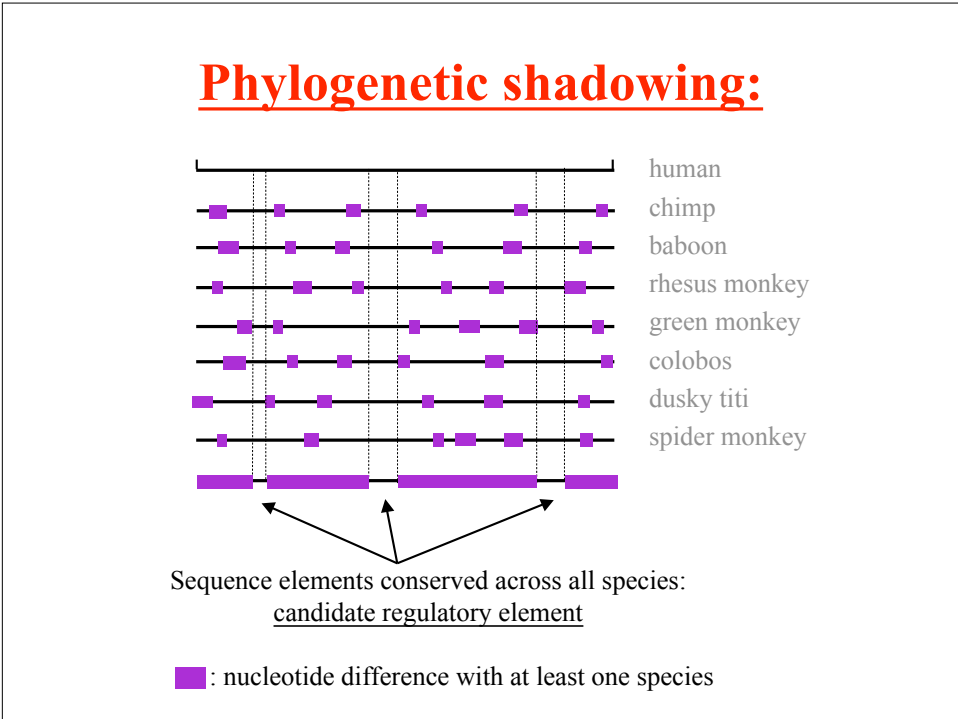
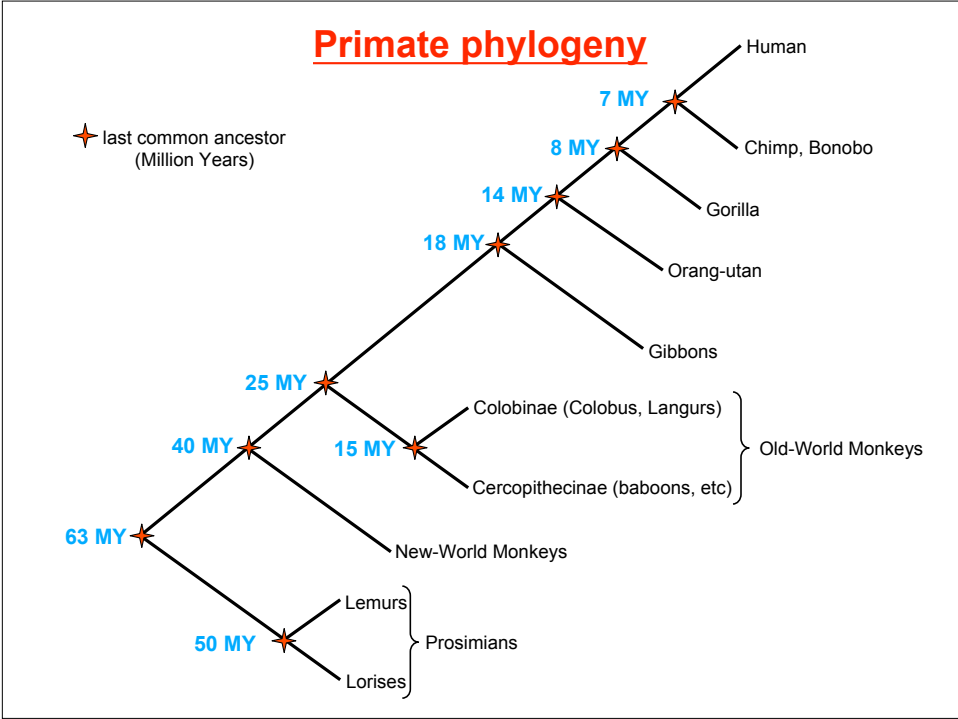


Pairwise species comparison



Approach to the Identification of Primate-Specific Regulatory Elements





1. PROOF-OF-PRINCIPLE:

Identification of intron-exon boundaries
of 4 genes with known boundaries

2. TEST-CASE:

Identification of regulatory elements of a
primate-specific gene

Amplification of target sequence from genomic DNA of ~20 primates
(new- and old- world monkeys, hominoids)



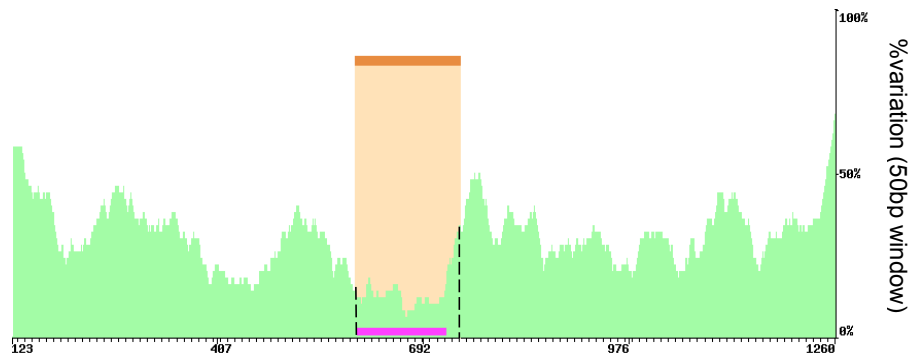
Direct sequence of the PCR product



Multiple sequence alignment with ClustalW:

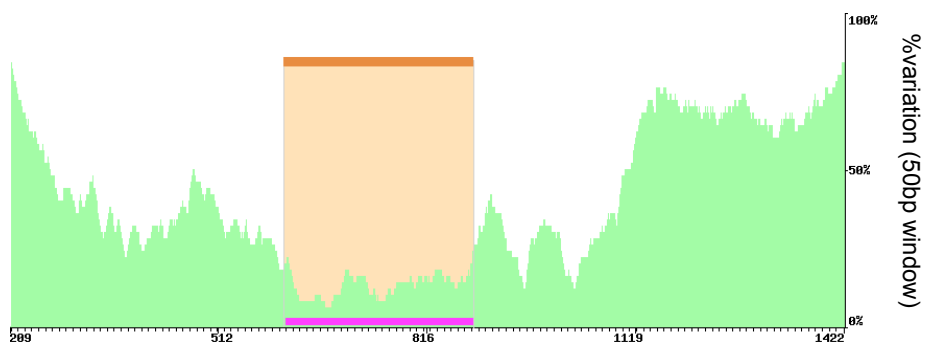
```
Allen          CACCCAGNTAATTTTGTATTTTAGTAGAG---ACAGGGTTTCACTACGTTGGCCAGGC 264
Green          CACCCGGNTAATTTTGTATTTTAGTAGAG---ACAGGGTTTCACTACGTTGGCCAGGC 193
Human          CACCCGGCTAATTTTGTATTTTAGTAGAG---TCGGGGTTTCACTATGTTGGCCAGGC 398
Chimp          CACCCGGCTAATTTTGTATTTTAGTAGAG---TCGGGGTTTCACTATGTTGGCCAGGC 387
Orangutan      CGCCCGGCTAATTTTGTATTTTAGTAGAG---TCGGGGTTTCACTATGTTGGCCAGGC 310
Colobus        CACCCGATAAATTTTGTATTTTAGTAGAG---ACGGGGTTTCACTACGTTGGCCAGGC 331
Douc           CACCCGGCTAATTTTGTATTTTAGTAGAG---ACGGGGTTTCACTACGTTGGCCAGGC 328
Francois       CACCCGGCTAATTTTGTATTTTAGTAGAG---ACGGGGTTTCACTACGTTGGCCAGGC 194
Drill          CACCCGGCTAATTTTGTATTTTAGTAGAG---ACAGGGTTTCACTACGTTGGCCAGGC 303
Mangabey       CACCCGNTAATTTTGTATTTTAGTAGAG---ACAGGGTTTCACTACGTTGGCCAGGC 192
Owl            CATCCGGCTAATTTTGTATTTTAGTAAGGCGACAGGATTCACCATGTTGGCCAGGC 311
Squirrel       CACCCGGCTAATTTTGTATTTTAGTAGAGGCGACAGGATTCACCATGTTGGCCAGGC 196
Tamarin        CACCTAATAA-TTTTGTATTTTAGAAAAGGCGACAGGATTCACCACGTTGGCCAGGC 331
Titi           CACCTGGCTAATTTTGTATTTTAGTACAGGCAACAGGATTCACCATGTTGGCCAGGC 322
Consensus      * *   *** ***** * **   * ** * * * * * * * * * * * * * * * * *
```

Identification of plasminogen exon 6



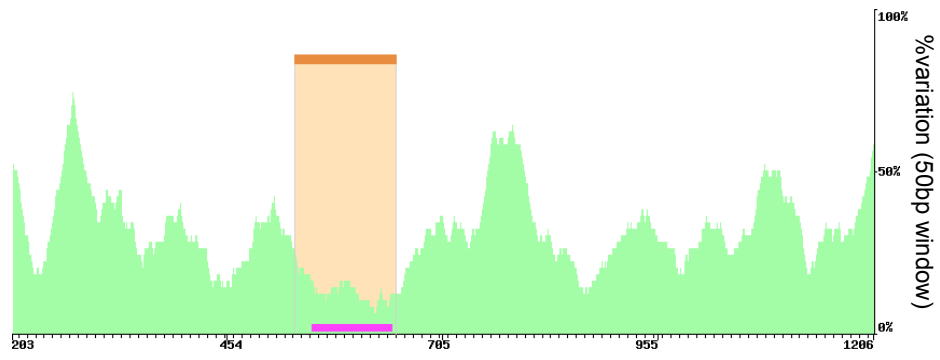
Exon prediction with a two-state Hidden Markov Model:
1) site averagely conserved
2) site very well conserved

Identification of LXR- α exon 3



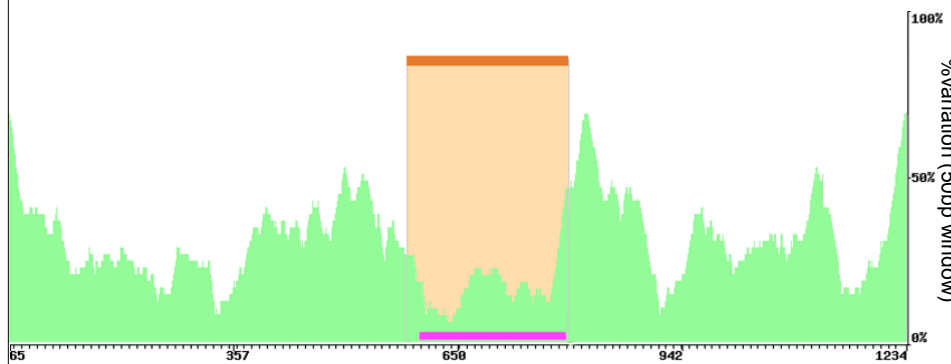
Exon prediction with a two-state Hidden Markov Model

Identification of CETP exon 8



Exon prediction with a two-state Hidden Markov Model

Identification of ApoB exon 19



Exon prediction with a two-state Hidden Markov Model

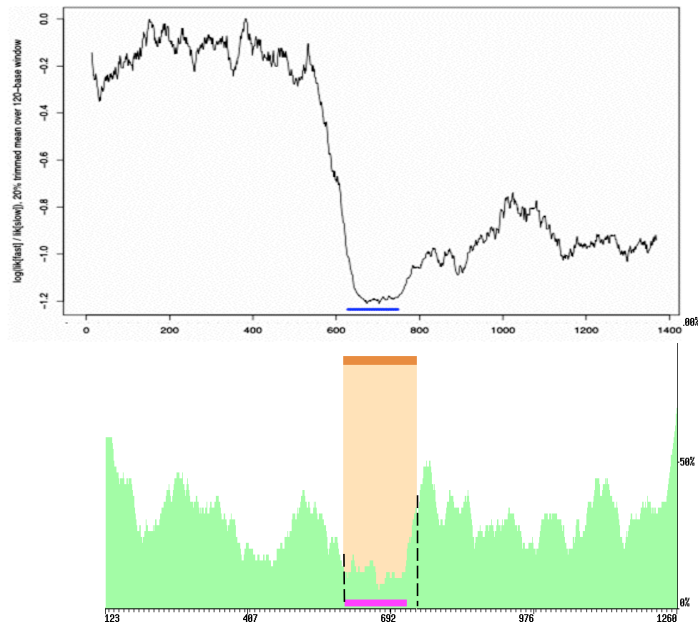


MAXIMUM LIKELIHOOD PHYLOGENETIC SHADOWING:

Learn mutation rates for 'conserved' and 'non-conserved' regions on a training set

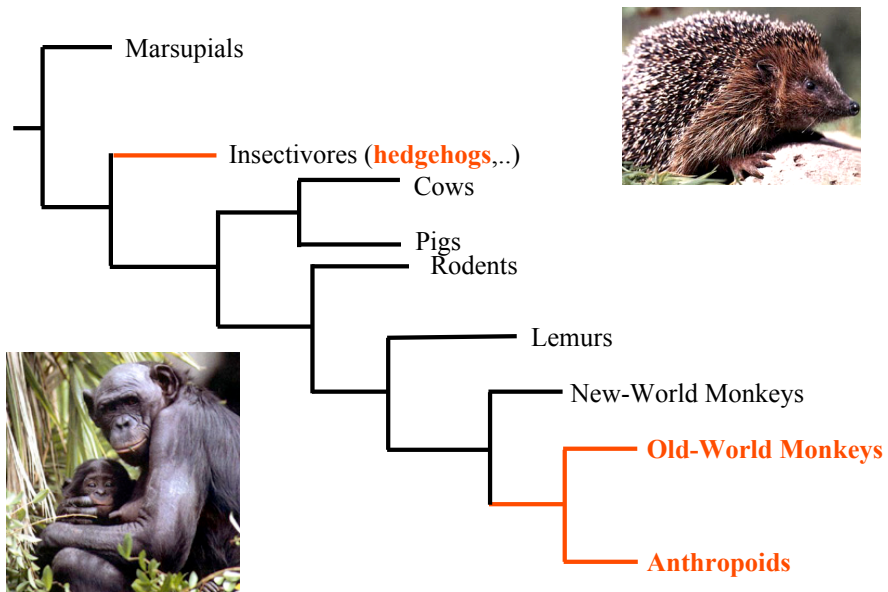
Calculate log-likelihood ratios of rates for each column of the multiple alignment using the phylogenetic trees.

Plasminogen exon 6



Amplification of evolutionary distance through multiple species comparison successfully identifies known functional elements

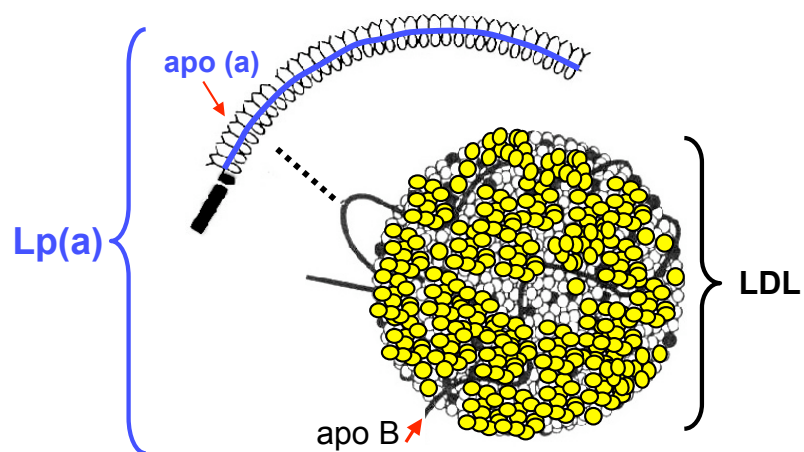
Identification of regulatory elements of a primate-specific gene: apo(a)



apolipoprotein(a)

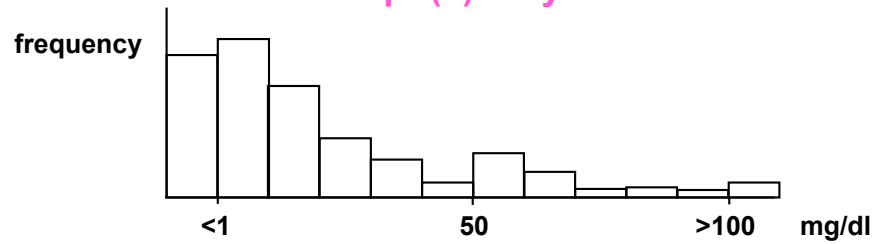
- Lacks a mouse ortholog:
- People with high levels are at risk of developing atherosclerosis

apo (a) is a plasma protein



apo(a) plasma levels are highly variable

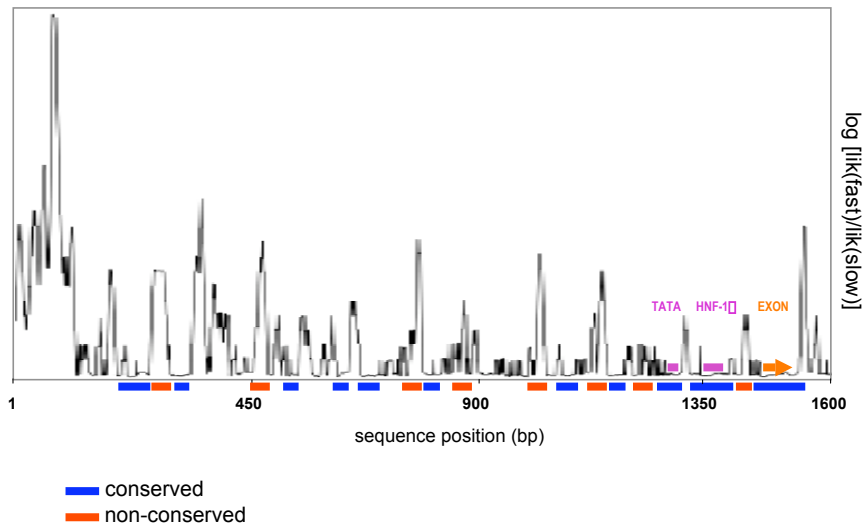
* apo(a) vary 1000-fold



* standard protein ~10-fold

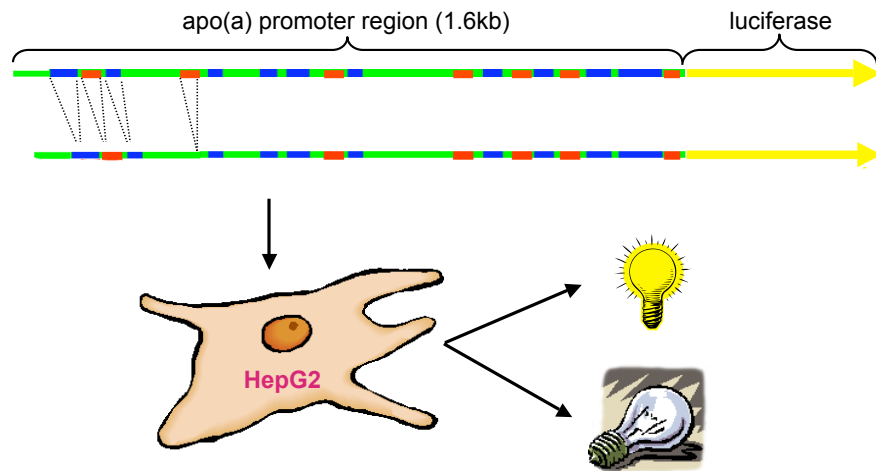
~ 50% of apo(a) variability in levels is attributable to the regulation of the transcription of the gene

Phylogenetic shadowing of the apo(a) promoter

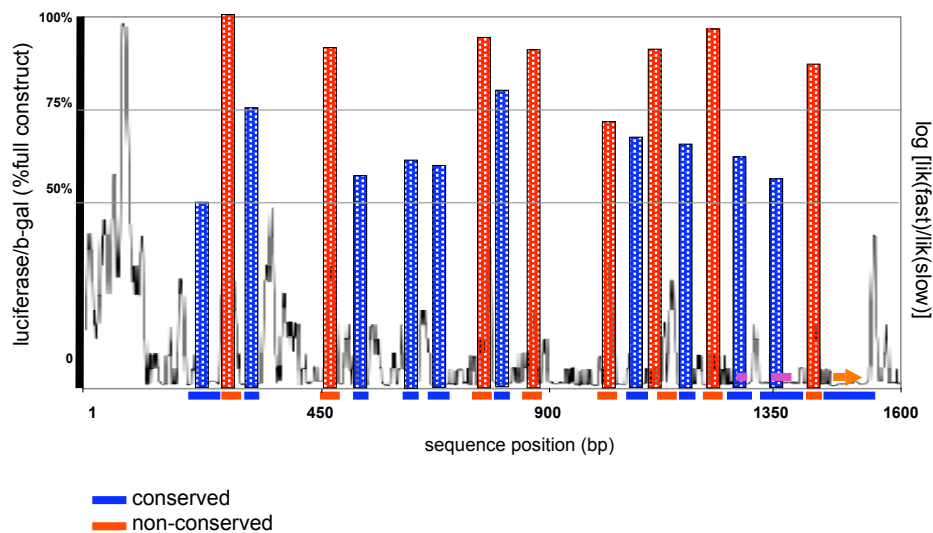


Testing the function of conserved elements

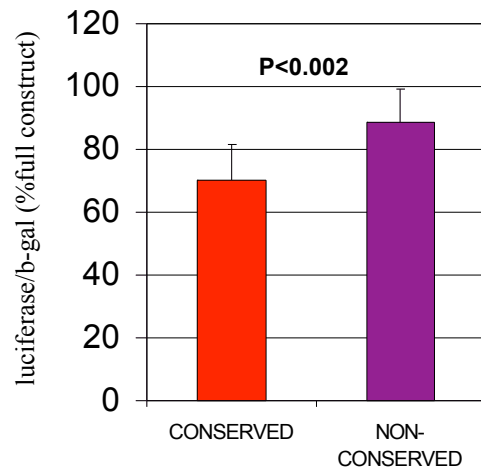
Cell culture transcription assay:



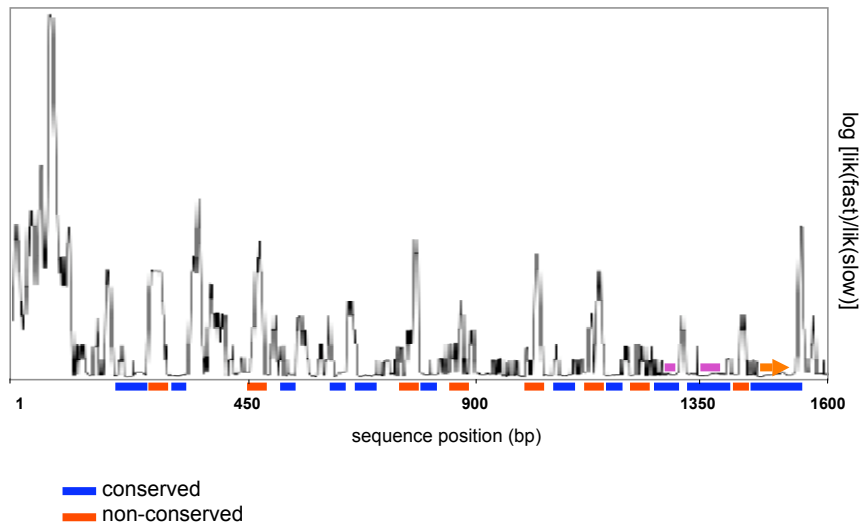
Phylogenetic shadowing of the apo(a) promoter



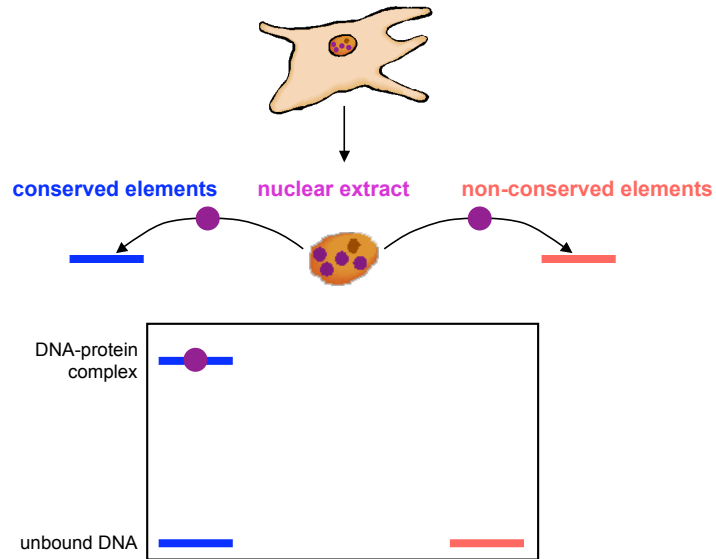
Average of expression data for conserved and non-conserved elements



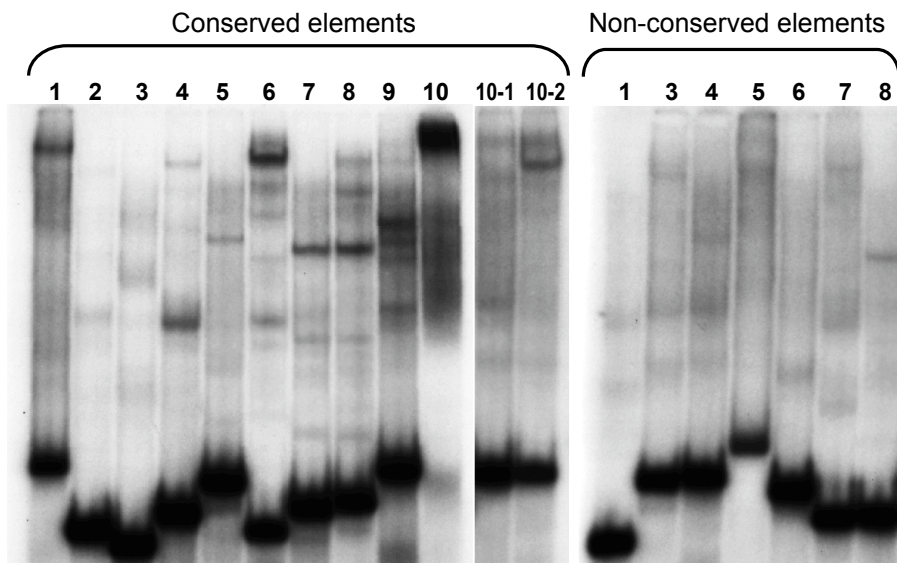
Phylogenetic shadowing of the apo(a) promoter



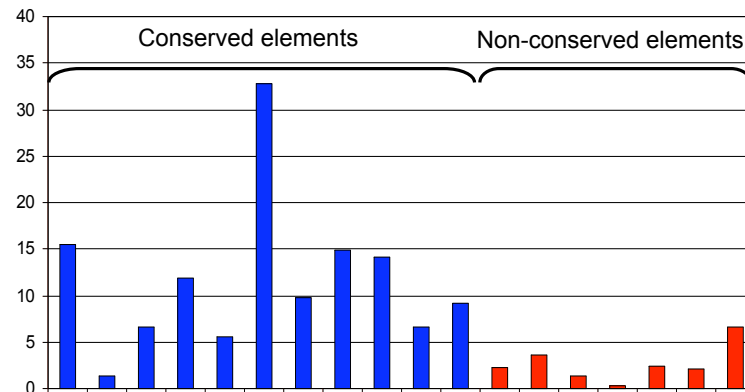
Gel-shift assay to assess DNA-protein interactions



Gel-shift analysis of conserved elements in the apo(a) promoter

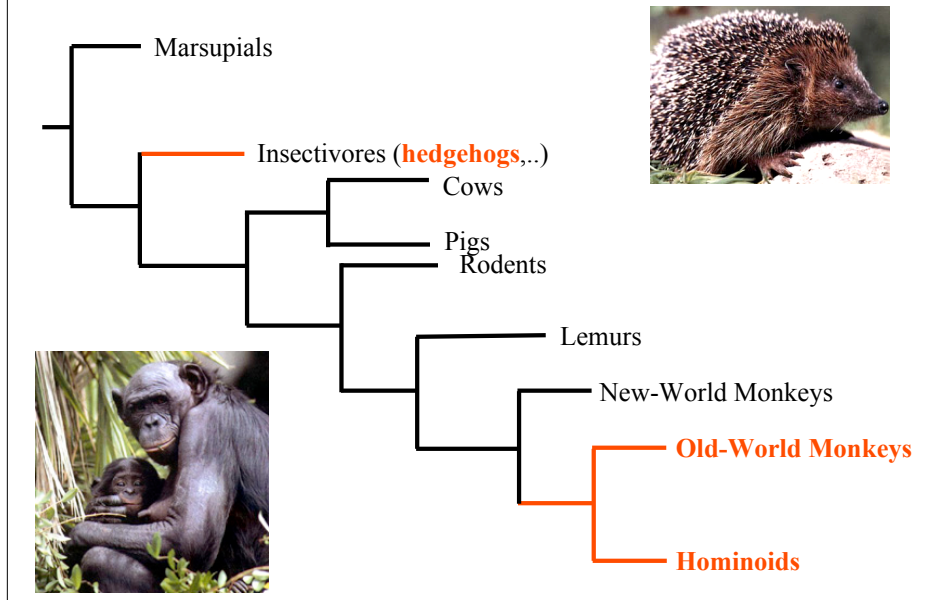


Quantitative analysis of the gel-shift pattern



genomic sequence from multiple primates can be used to identify functional sequence shared by this group of species

apo(a) is a recently evolved gene but...



Strategy to unravel the mechanism of evolution of the apo(a) gene

Sequence the genomic region containing apo(a) and its orthologous region in species without apo(a):

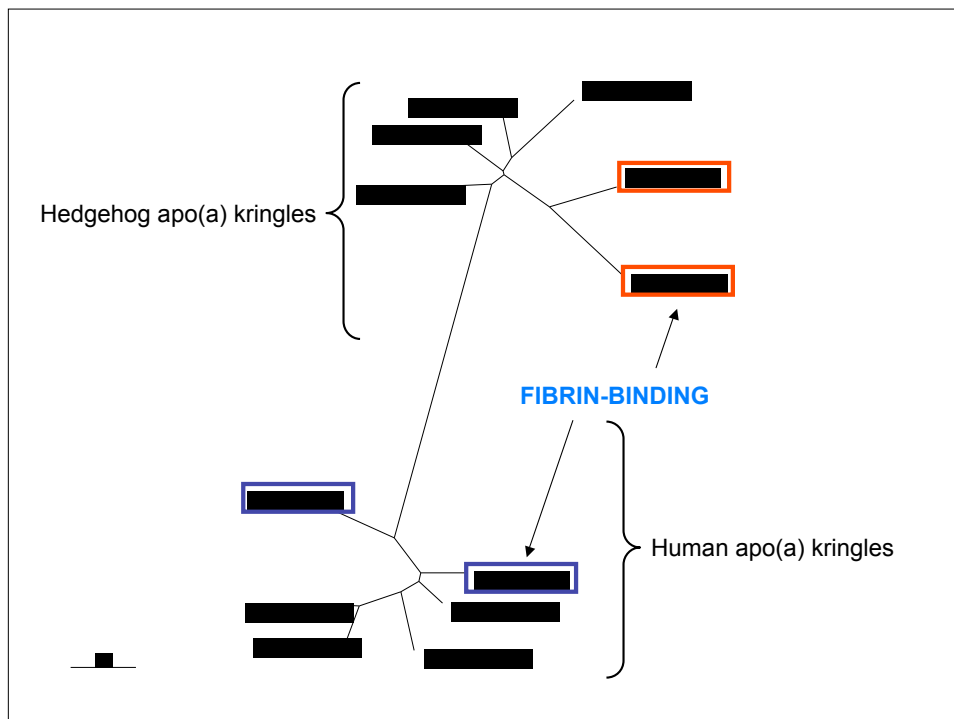
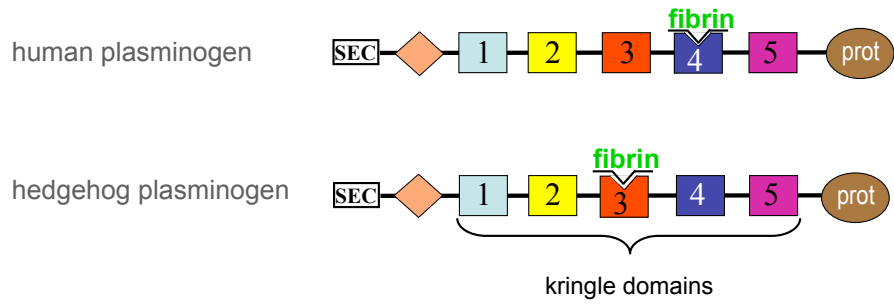
Species with apo(a):

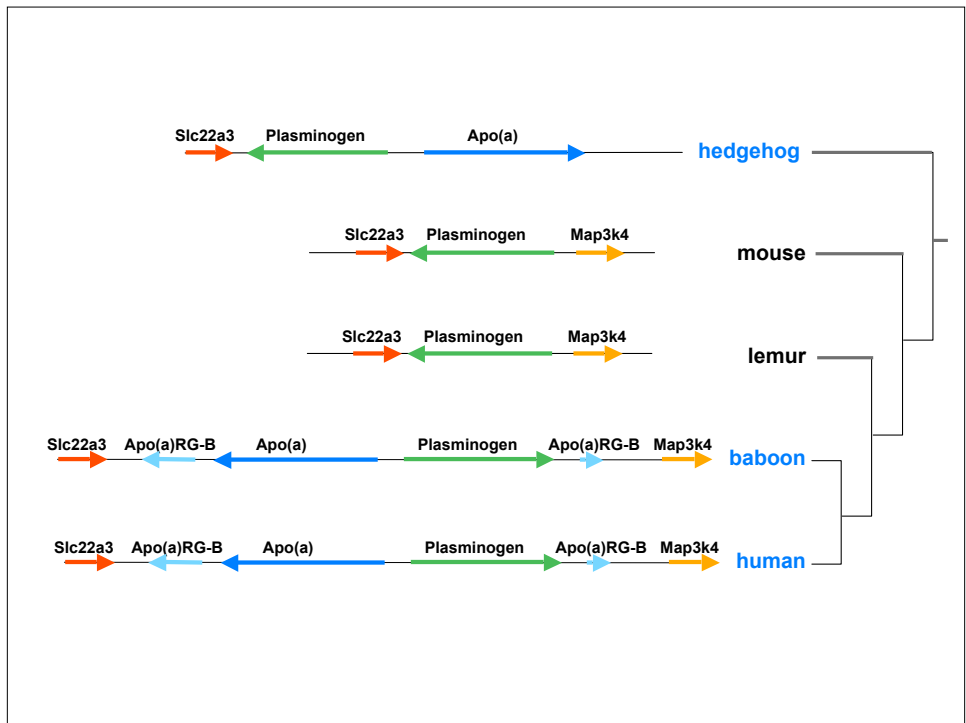
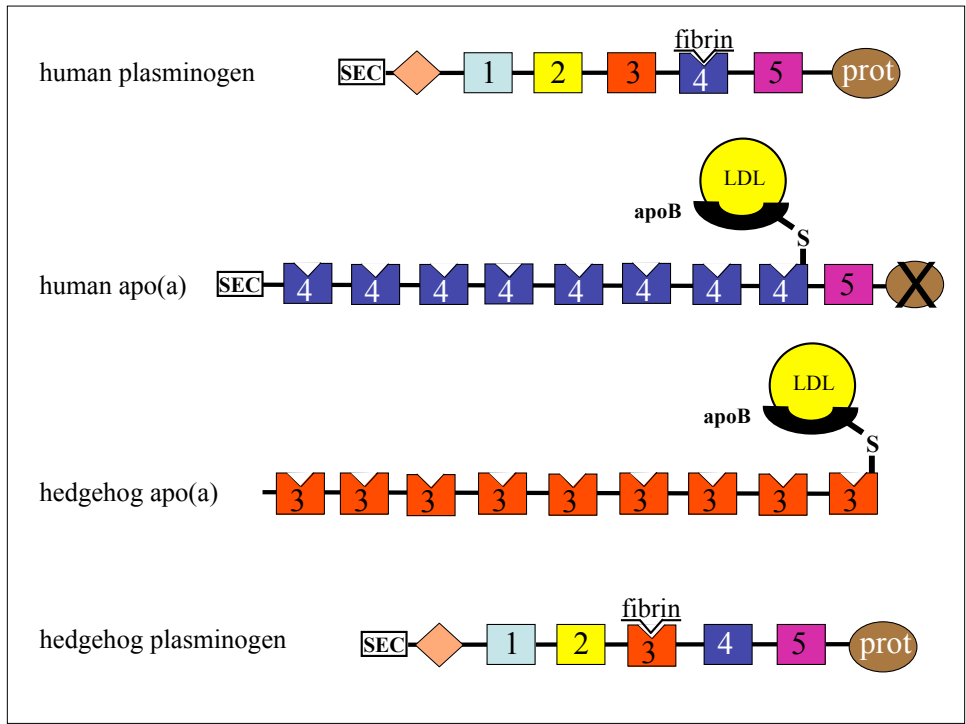
- human: publicly available
- baboon: sequence
- hedgehog: generate BAC library and sequence

Species without apo(a):

- mouse: sequence
- lemur: generate BAC library and sequence

human apo(a) is a duplication of plasminogen





apo(a) is a unique example of convergent evolution where the same gene arose twice through the independent duplication and remodeling of the same paralogous gene